

PATENT
450100-03197

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION FOR LETTERS PATENT

13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
222

DATA TRANSMISSION METHOD, APPARATUS USING SAME, AND DATA
TRANSMISSION SYSTEM

5 BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to a method for transmitting data including images during communication and streaming among a plurality of terminals, a data transmission apparatus such as a mobile phone, a PDA, a TV telephone terminal, and a PC terminal, and a data transmission system.

2. Description of the Related Art

As a communication service for communications and streaming among a plurality of terminals, there is for example, as illustrated in Fig. 1, the visual communication service for transmission of data and images by mobile terminals (MT) comprised of portable phones.

The visual communication service system 1 compresses voice and images and exchanges signals between a mobile terminal (hereinafter simply referred to as a "terminal") 2-1 and terminal 2-2 through a transmission channel 3 to enable conversation to be enjoyed while viewing the other party's face.

Each of the terminals 2-1 and 2-2, whose

appearances are illustrated in Fig. 1, has a body case 21, an operation panel 22 provided at the lower portion of the body case 21 and including a keypad for inputting a telephone number for a call function or for inputting certain operational instructions, a display monitor 23 provided above the operational panel 22, a camera 24 provided above the display monitor 23 (at the upper portion of the body case 21), and an antenna 25.

The transmission channel 3 includes mobile base stations (MBS) 31-1 and 31-2, mobile switching centers (MSC) 32-1 and 32-2, gateway mobile switching centers (GMSC) 33-1 and 33-2 having home location registers (HLR), and a circuit network 35.

The mobile base stations 31-1 and 31-2 communicate with the terminals 2-1 and 2-2. The signals handled by the base stations 31-1 and 31-2 are sent to the higher level mobile switching centers 32-1 and 32-2 and further to a backbone circuit network 35 through the higher level gateway mobile switching centers 33-1 and 33-2.

When transmitting images to such a circuit network 35, however, the information carrying capacity of the circuit network 35 is limited by the bandwidth of the circuit network 35.

Therefore, a user who is talking using for

example the terminal 2-1 can view his or her own image captured by the camera 24 at a high quality such as with the image shown by 23a in Fig. 1 because the image is monitored directly, but after compression at the terminal 5 2-1 and transmission to the terminal 2-2, the quality of the image sent from the terminal 2-1 becomes extremely degraded at the terminal 2-2 such as shown by the image shown by 23b.

Similarly, at the terminal 2-1, the quality of 10 the image sent from the terminal 2-2 to the terminal 2-1 becomes extremely degraded.

In this way, in an image transmission system of the prior art, the transmitted images become extremely degraded. The difference from the image which the user 15 himself or herself monitors becomes too great. This causes not only a strong uncomfortable feeling but also a feeling of stress as the expression of the other party cannot be seen clearly.

Further, in each of the terminals 2-1 or 2-2, 20 the camera 24 is located above the display monitor 23 of the terminal. The users look at their display monitors 23 of their terminals. Therefore there is the problem that both end up looking down and therefore conversation maintaining eye contact becomes impossible.

SUMMARY OF THE INVENTION

An object of the present invention is to provide a data transmission method able to maintain the quality of a transmitted image substantially the same as that of one's own image and an apparatus using the method and a data transmission system.

A second object of the present invention is to provide a data transmission method able to realize communication in a state where eye contact is maintained and an apparatus using the method and a data transmission system.

In order to achieve the above objects, the present invention provides a data transmission method for transmitting at least image data accommodated in an image frame of a preset size, comprising the steps of capturing an image including a target image to be transmitted; making adjustments so that the target image to be transmitted from the captured image substantially fully fills the image frame; and compressing and transmitting the adjusted image data in the image frame.

Further, the present invention provides a data transmission method for transmitting at least image data comprising the steps of capturing an image including a target image to be transmitted; making adjustments so that the target image to be transmitted from the captured

image has a preset size; and compressing and transmitting the adjusted image data in the image frame.

The method of the present invention further comprises, when making adjustments so that a target image
5 fully fills an image frame or when making adjustments so that the target image to be transmitted from the captured image has a preset size, using an image of a characteristic portion of the image to be transmitted for matching and cutting out the image around the part with
10 the best match for tracking of the image.

Alternatively, the method of the present invention further comprises, when making adjustments so that a target image fully fills an image frame or when making adjustments so that the target image to be transmitted
15 from the captured image has a preset size, using an image of a characteristic portion of the image to be transmitted to calculate a distance in the characteristic portion and using this for enlargement, reduction, and tracking of the image.

20 Alternatively, the method of the present invention further comprises, when detecting a characteristic portion of an image, locking an image in a desired image state while monitoring a state of input of the image to be transmitted and using a center portion of the locked
25 image as the characteristic portion of the image.

Alternatively, the method of the present invention further comprises, when detecting a characteristic portion of an image, displaying a state of input of the image to be transmitted, specifying a certain location on the display screen, and using the portion around the specified point as the characteristic portion of the image.

Alternatively, the method of the present invention further comprises, when locking a target image by using a characteristic portion of an image, fetching and tracking the image at a frame rate the same as or higher than the image to be transmitted and refreshing a reference image serving as a reference for tracking at a rate the same as or higher than the transmission frame rate.

Alternatively, the method of the present invention further comprises displaying a state of input of a transmitted image and specifying a range to be transmitted on the display to determine the range of transmission of the image.

Further, the present invention provides a data transmission method for displaying at least image data of a user on a display screen among a plurality of terminals and transmitting at least image data while displaying received image data on the display screen, comprising capturing an image including a target image of a user to

be transmitted from substantially the center of the display screen.

Further, the present invention provides a data transmission method for displaying at least image data of
5 user on a display screen among a plurality of terminals and transmitting at least image data while displaying received image data of another party on the display screen and displaying content for common discussion on the display screen, comprising the steps of capturing an
10 image including a target image of a user to be transmitted from substantially a center of the display screen and displaying the image data at one of an upper portion or a lower portion of the display screen from the substantially center portion of the display screen
15 serving as the capturing portion and displaying the content at the other portion.

That is, the method of the present invention comprises displaying image data above the content on the display screen or displaying image data below the content
20 on the display screen.

Further, the present invention provides a data transmission apparatus for transmitting at least image data accommodated in an image frame of a preset size, comprising an imaging means for capturing an image
25 including a target image to be transmitted; a first

circuit for making adjustments so that the target image to be transmitted from the captured image by the imaging means substantially fully fills the image frame; and a second circuit for compressing and transmitting the adjusted image data in the image frame.

Further, the present invention provides a data transmission apparatus for transmitting at least image data, comprising an imaging means for capturing an image including a target image to be transmitted; a first circuit for making adjustments so that the target image to be transmitted from the captured image has a preset size; and a second circuit for compressing and transmitting the adjusted image data in the image frame.

In the data transmission apparatus of the present invention, when making adjustments so that a target image fully fills an image frame or when making adjustments so that the target image to be transmitted has a preset size, the first circuit uses an image of a characteristic portion of the image to be transmitted for matching and cuts out the image around the part with the best match for tracking of the image.

Alternatively, in the data transmission apparatus of the present invention, when making adjustments so that a target image fully fills an image frame or when making adjustments so that the target image to be transmitted

has a preset size, the first circuit uses an image of a characteristic portion of the image to be transmitted to calculate a distance in the characteristic portion and uses this for enlargement, reduction, and tracking of the
5 image.

Alternatively, the data transmission apparatus of the present invention comprises a displaying means for displaying image data and a locking means able to lock an image displayed by the displaying means in a desired
10 state, wherein when detecting a characteristic portion of an image, the first circuit uses a center portion of the image locked by the locking image as the characteristic portion of the image.

Alternatively, the data transmission apparatus of
15 the present invention comprises a displaying means for displaying image data and including a pointer able to specify a certain location and a specifying means able to specify a certain location of the image displayed on the displaying means by the pointer, wherein when detecting a
20 characteristic portion of an image, the first circuit uses a portion around the point specified by the specifying means as the characteristic portion of the image.

Alternatively, in the data transmission apparatus of
25 the present invention, when locking a target image by

using a characteristic portion of the image, the first circuit fetches the image and performs a tracking operation at a frame rate the same as or higher than the image to be transmitted and refreshes a reference image
5 serving as a reference for tracking at a rate the same as or higher than the transmission frame rate.

Alternatively, the data transmission apparatus of the present invention comprises a displaying means for displaying image data and including a pointer able to
10 specify a certain location and a specifying means able to specify a certain location of the image displayed by the displaying means by the pointer, wherein the first circuit decides on the range specified by the specifying means as the range for transmission of the image.

15 Further, the present invention provides a data transmission system for displaying at least image data of a user on a display screen among a plurality of terminals and transmitting at least image data while displaying received image data of another party on the display
20 screen, wherein a terminal comprises an imaging means for capturing an image including a target image to be transmitted; a first circuit for making adjustments so that the target image to be transmitted from the captured image by the imaging means substantially fully fills an
25 image frame; and a second circuit for compressing and

transmitting the adjusted image data in the image frame.

Further, the present invention provides a data transmission system for displaying at least image data of a user on a display screen among a plurality of terminals and transmitting at least image data while displaying
5 received image data of another party on the display screen, wherein a terminal comprises an imaging means for capturing an image including a target image to be transmitted; a first circuit for making adjustments so
10 that the target image to be transmitted from the captured image becomes a preset size; and a second circuit for compressing and transmitting the adjusted image data in an image frame.

Further, the present invention provides a data
15 transmission system for displaying at least image data of a user on a display screen among a plurality of terminals and for transmitting at least image data while displaying received image data on the display screen, wherein an
imaging means for capturing an image including a target
20 image to be transmitted is provided substantially at the center of the display screen.

Further, the present invention provides a data transmission system for displaying at least image data of a user on a display screen among a plurality of terminals
25 and for transmitting at least image data while displaying

received image data of another party on the display screen and displaying content for common discussion on the display screen comprising an imaging means provided substantially at the center of the display screen and

5 capturing an image including a target image to be transmitted and a displaying means for displaying the image data at one of an upper portion or a lower portion of the display screen from the substantially center portion of the display screen serving as the capturing

10 portion and displaying the content at the other portion.

According to the present invention, in order for a target image such as an image of a user's face to fully fill an image frame, the face image is enlarged, reduced, or tracked according to its movement. Then, only the face

15 image is cut out and that image is compressed and transmitted.

Further, according to the present invention, in order for the size of a user's transmitted image to be kept constantly the same, the image is enlarged, reduced,

20 or tracked according to its movement. Due to this, an image of a person of the same size is cut out and that image is compressed and transmitted.

Further, by configuring the invention so that a user is able to lock (keep) an image in a desired state while

25 viewing the monitor screen, the image can be tracked,

enlarged, reduced, or cut out on the terminal according to the instructions of the user.

Further, in order to continue the locked state, at the time of starting locking, for example, a
5 characteristic portion (a portion of the body such as the eyes, nose, mouth, ears, or outline of the face) is cut out from the image. Using this portion as a reference, the image is enlarged, reduced, or tracked.

For example, the distance between the two eyes is
10 detected, and from this distance, the rate of enlargement or reduction or the position of cut out is detected.

In addition, the entire image input from the imaging means, that is, the camera, is output for the image on the monitor of the user. The user specifies a region of
15 the image to cut out that region. That image is then reduced, enlarged, or tracked so that this image subsequently remains the same size.

Further, by using an operable pointer on a monitor screen, a user specifies a characteristic image at the
20 time of starting locking and therefore that image is reduced, enlarged, or tracked.

Further, in order to increase the tracking accuracy, the image is fetched and tracked at a frame rate the same as or higher than the image to be transmitted, and a
25 reference image serving as a reference for tracking is

refreshed at a rate the same as or higher than the transmission frame rate.

Further, an imaging means is embedded in the display screen. Therefore, the lines of sight of the mutually transmitted images can be matched.

In addition, according to the present invention, for example, a face image and content for discussion are provided above and below each other and an imaging means is provided near the place between them, so images of users talking while viewing the image under discussion can be transmitted to each other. Further, virtual reality communication where eye contact can be maintained when viewing each others' faces on the monitor can be realized.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other objects and features of the present invention will become clearer from the following description of the preferred embodiments given with reference to the attached drawings, in which:

Fig. 1 is a view of terminals and a transmission channel in visual communication of the related art;

Fig. 2 is a view of the system configuration of an embodiment of a visual communication service system as a data transmission system according to the present

invention;

Fig. 3 is a view of the appearance showing one example of the configuration of a mobile terminal according to the present embodiment;

5 Fig. 4 is a circuit diagram of a first embodiment of a signal processing system of a terminal as a data transmission apparatus according to the present invention;

Fig. 5 is a circuit diagram of a pre-processing circuit in Fig. 4 according to the first embodiment.

Figs. 6A and 6B are schematic views for explaining the tracking in the pre-processing circuit;

Figs. 7A to 7C are views for explaining a first method for determination of an initial reference image;

15 Figs. 8A to 8C are views for explaining a second method for determination of an initial reference image;

Fig. 9 is a circuit diagram of a second embodiment of a signal processing system of a terminal as a data transmission apparatus according to the present invention;

Fig. 10 is a circuit diagram of a pre-processing circuit in Fig. 9 according to the second embodiment;

Fig. 11 is a view for explaining an example of calculating a face size;

25 Fig. 12 is a view for explaining an example of a

face part detection search region;

Fig. 13 is a view for explaining processing for reducing a face image in a face part detection search region and displaying the same on a display monitor;

5 Fig. 14 is a view for explaining processing for reducing a face image in a face part detection search region and displaying the same on a display monitor;

Fig. 15 is a view for explaining processing for enlarging a face image in a face part detection search region and displaying the same on a display monitor;

Fig. 16 is a view for explaining processing for enlarging a face image in a face part detection search region and displaying the same on a display monitor;

Fig. 17 is a circuit diagram of a third embodiment of a signal processing system of a terminal as a data transmission apparatus according to the present invention;

Fig. 18 is a circuit diagram of a pre-processing circuit in Fig. 17 according to the third embodiment;

20 Fig. 19 is a view for explaining processing for reducing a face image in a specified an image transmission region and displaying it on a display monitor according to the third embodiment;

Fig. 20 is a view for explaining processing for reducing a face image in a specified an image

25

transmission region and displaying it on a display monitor according to the third embodiment;

Fig. 21 is a view for explaining a ratio of pixels "e" of a face color over pixels "f" in an image on a monitor according to the third embodiment;

Fig. 22 is a view for explaining processing for reducing a face image in a face part detection search region and displaying it on a display monitor according to the third embodiment;

Fig. 23 is a view for explaining processing for reducing a face image in a face part detection search region and displaying it on a display monitor according to the third embodiment; and

Fig. 24 is a conceptual view for explaining processing for fetching and tracking an image at a frame rate the same as or higher than a transmitted image and refreshing a reference image for tracking at a rate the same as or higher than the transmission frame rate in order to increase a tracking accuracy of the image according to the fourth embodiment.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Below, preferred embodiments of the present invention will be explained with reference to Fig. 1 to

Figure 2 is a view of the configuration of an embodiment of a visual communication service system as a data transmission system according to the present invention.

5 The visual communication service system 100 compresses voice and image data and exchanges signals among a plurality of terminals, for example, between a mobile terminal (hereinafter simply referred to as a "terminal") 50-1 and terminal 50-2 through a transmission
10 channel 60. By this system, a user is able to enjoy conversation with another party while viewing his or her face.

 Figure 3 is a view of the appearance showing an example of the configuration of the terminal 50-1 or 50-
15 2.

 The terminal 50-1 or 50-2, whose appearance is illustrated in Fig. 3, has a body case 51, an operation panel 52 provided at the lower portion of the body case 51 and including a keypad for inputting a telephone
20 number for a call function or for inputting specific operational instructions, a display monitor 53 provided above the operational panel 52, a camera module 54 including a lens arranged at substantially a center region of the display monitor 53 and serving as an
25 imaging means, an antenna 55, and a switch 56 acting as a

locking means or specifying means and provided with functions as a GUI.

In Fig. 3, a microphone and a speaker are omitted. The specific circuit configuration of a terminal acting
5 as a data transmission apparatus will be described in detail later.

In this way, the display monitor 53 has a lens portion of the camera module 54 arranged at the substantially center region. As shown in Fig. 3, face
10 images and a predetermined content image obtained from a later explained application service provider (ASP) are displayed above and below the position of the lens portion.

In the example in Fig. 3, the face images are shown
15 at the upper portion, and the contents image is shown at the lower portion, but it is also possible to conversely display the content images at the upper portion and the face images at the lower portion.

The transmission channel 60 includes mobile base
20 stations (MBS) 61-1 and 61-2, mobile switching centers (MSC) 62-1 and 62-2, gateway mobile switching centers (GMSC) 63-1 and 63-2 having home location registers (HLR), and a circuit network 64.

The mobile base stations 61-1 and 61-2 communicates
25 with the terminals 50-1 and 50-2, respectively. The

signals handled by the base stations 61-1 and 61-2 are transmitted to the higher level mobile switching center 62-1 and 62-2 and further to a backbone circuit network 64 through the higher level gateway mobile switching centers 63-1 and 63-2.

Further, a gateway (GW) 65 is connected to the circuit network 64, while a application service provider 66 able to distribute content is connected with the gateway 65.

Therefore, it is possible to start up the application service provider (ASP) 66 offering various content CONT and output the same images of the content CONT to the terminals 50-1 and 50-2 through the gateway 65.

Below, circuit configurations of specific signal processing systems in terminals 50 (terminals 50-1 and 50-2) acting as data transmission apparatuses according to the present invention, as applied to the above visual communication service system 100, will be explained as first, second, and third embodiments with reference to the accompanying drawings.

First Embodiment

Figure 4 is a circuit diagram of a first embodiment of a signal processing system of a terminal functioning as a data transmission apparatus according to the present

invention.

The signal processing system 500, as shown in Fig. 4, has a microphone 501, a camera circuit 502, a pre-processing circuit 503 as the first circuit, a video
5 compressor and coder 504, a voice compressor and coder 505, a multiplexer (MUX) 506, a transmission circuit 507, a reception circuit 508, a demultiplexer (DeMUX) 509, a video decoding circuit 510, a voice decoding circuit 511, an image combining apparatus 512, a display circuit 513,
10 a speaker 514, a GUI interface (I/F) circuit 515, and a CPU 516.

The video compressor and coder 504, voice compressor and coder 505, multiplexer (MUX) 506, and transmission circuit 507 form the second circuit.

15 The microphone 501 converts the voice generated by a user into an electrical signal and outputs it to the voice compressor and coder 505.

The camera circuit 502 performs certain processing on an image of for example the upper part of a user's
20 body captured by the camera module 54 through its lens and converted from an optical signal into an electrical signal and outputs the signal as a digital image signal S502 to the pre-processing circuit 503.

The pre-processing circuit 503 receives the digital
25 image signal S502, extracts an image to be transmitted

(hereinafter referred to as a "target image") from the image of the upper part of the user's body, generates an image signal adjusted so that the image to be transmitted fully fills the image frame, and outputs it to the video
5 compressor and coder 504 and image combining circuit 512.

As the processing for adjusting the target image to fully fill the image frame, the pre-processing circuit 503 for example uses the image of a characteristic portion of the image to be transmitted for matching and
10 cuts out the images around the part of the best match so as thereby to enable tracking of the image.

Here, "fully fill the image frame" means to thoroughly fill it with the target image until the background is not recognizable. Especially, in image
15 transmission apparatuses for communication, the target images are mostly face images.

Below, the processing of the pre-processing circuit 503 will be described in further detail.

Figure 5 is a circuit diagram of an example of the
20 specific configuration of a principal portion of the pre-processing circuit according to the first embodiment.

In addition, Figs. 6A and 6B are schematic views for explaining the tracking processing by the pre-processing circuit.

25 The pre-processing circuit 503, as shown in Fig. 5,

comprises a search region extractor 5031, a binary-coded unit 5032, a pattern matching unit 5033, a reference memory 5034, and a rectangular region extractor 5035.

The search region extractor 5031 extracts a search
5 region SRG as shown in Fig. 6B based on the coordinate information of the reference image RIM of the previous image PIM shown in Fig. 6A and stored in the reference memory 5034.

The binary-coded unit converter 5032 converts the
10 luminance signal of the search region data extracted by the search region extractor 5031 into binary data and outputs it to the pattern matching unit 5033.

The pattern matching unit 5033 performs pattern
15 matching between the luminance information of the binary data of the luminance signal of the search region given by the binary-coded unit 5032 and of the binary data of the reference image RIM of the previous image PIM shown in Fig. 6A and stored in the reference memory 5034 and outputs the results to the rectangular region extractor
20 5035.

The pattern matching processing employs the image data MMT at a position giving the highest consistency in the search region SRG as shown in Fig. 6B and uses it as the tracking results of the current image CIM.

25 In addition, the pattern matching unit 5033 stores

the tracking results at the present stage, that is, the image data MMT at a position giving the highest consistency, into the reference memory 5034 as a reference image for the next image.

5 The rectangular region extractor 5033 cuts out a certain rectangular region surrounding the tracking results MMT obtained by the pattern matching unit 5033, and outputs the same.

10 In the pre-processing circuit 503 having such a configuration, tracking of the input image is performed in order to maintain a desired image state (hereinafter referred to as an "initial reference image") that a user wishes to transmit.

15 Then, the tracking operation is performed at a frame rate the same as or higher than the transmission frame rate, and a reference image serving as a reference for tracking is refreshed successively.

20 The tracking in the pre-processing circuit 503 is started when information of an initial reference image (described later with reference to Figs. 7A to 7C and Figs. 8A to 8C) is transmitted from the CPU 516.

25 First, in the input image data (the current image) CIM newly input from the camera 502 as shown in Fig. 6B, a search region SRG is extracted by the search region extractor 5031 based on the coordinate information of the

reference image RIM of the previous image PIM stored in the reference memory 5034.

Next, the luminance signal of the extracted data is converted into binary data in the binary-coded unit 5033, then pattern matching of luminance information is performed by the pattern matching unit 5033 between the luminance signal of the extracted data and the binary-coded data of the reference image RIM stored in the reference memory 5034.

The tracking results at this time, namely, the image data MMT at a position giving the highest consistency, is stored in the reference memory 5034 as a reference image of the next image.

Then, the rectangular region extractor 5035 cuts out and outputs a rectangular region surrounding the tracking result MMT obtained by the pattern matching unit 5033.

Next, first and second methods for determination of the initial reference image will be explained with reference to Figs. 7A to 7C and Figs. 8A to 8C.

The first method of determination determines the image by monitoring the image to be transmitted the image size frame.

In more detail, as shown in Fig. 7A, a transmission image size frame FR is set beforehand in the initial image IIM. The image in the frame FR is output as a

monitor image MNTa and, as shown in Fig. 7B, is displayed on the display monitor 53 by the displaying circuit 513 in Fig. 4.

The user views the output on the monitor and when a
5 desired image state is found, for example presses the switch 56 to issues a command for locking the image. Therefore, as shown in Fig. 7C, the initial reference image IRI is decided.

On the other hand, the second method of
10 determination decides on the image by monitoring a reduced image.

In more detail, as shown in Fig. 8A, the initial image IIM is reduced and output as a monitor image MNTb as shown in Fig. 8B. The reference center desired by the
15 user and the display region (the region occupied by the image) are specified using a pointer POI to decide the initial reference image IRI.

Note that, the instruction for moving the pointer POI on the display monitor 53 is executed by using for
20 example a ball switch, that is, the switch 56. Namely, by rotating the switch 56 in a desired direction, the pointer POI is moved up, down, left, right, or in an inclined direction.

Information on the initial reference image decided
25 by the first or the second method in this way is sent

from the GUI interface circuit 515 to the reference memory 5034 of the pre-processing circuit 503 through the CPU 516.

Further, in Fig. 4, the video compressor and coder
5 504 compresses and codes the transmitted image data extracted by the pre-processing circuit 503 and outputs it to the multiplexer 506.

The voice compressor and coder 505 compresses and codes the voice data from the microphone 501 and outputs
10 it to the multiplexer 506.

The multiplexer 506 multiplexes the image data to be transmitted, which was compressed and coded by the video compressor and coder 504, and the voice stream compressed and coded by the voice compressor and coder 505 and
15 outputs the data to the transmission circuit 507.

The transmission circuit 507 converts and modulates the image and voice data multiplexed by the multiplexer 506 into a specific transmission protocol and transmits it to the mobile base station (MBS) 61-1 or 61-2.

20 The reception circuit 508 receives and demodulates the transmitted signal sent from the other communication party through the transmission channel 60, decompresses the transmission protocol, and outputs the result to the demultiplexer 509.

25 The demultiplexer 509 separates the image data and

the voice data from the reception circuit 508 and supplies the obtained image data stream and the voice data stream to the video decoding circuit 510 and the voice decoding circuit 511, respectively.

5 The video decoding circuit 510 decodes the image data stream separated by the demultiplexer 509 and outputs it to the image combining circuit 512.

6
7
8
9
10 The voice decoding circuit 511 decodes the voice data stream separated by the demultiplexer 509 and outputs it to the speaker 514 for reproduction.

11
12
13
14 The image combining circuit 512 combines the face image data of a user of a terminal from the pre-processing circuit 503 and the face image data of for example the other party from the video decoding circuit
15 510 or a specified content image from the application service provider (ASP) 66 and outputs the result to the display circuit 513.

16
17
18 The display circuit 513 receives the combined image from the image combining circuit 512, displays on the
20 display monitor 53 of the terminal 50-1 (or 50-2), as shown in Fig. 3 and Fig. 4, a user's own face image and a transmitted face image of his other party for example side by side above the lens of the camera module 54, and displays the specified content image below the lens.

25 The GUI interface circuit 515 includes a switch 56.

The CPU 516 controls the entire terminal 50-1 (or 50-2) according to the instructions from the user to shift to the image lock state, switch the display on the monitor, and so on.

5 Next, the operation of the terminal 50 of Fig. 4 will be described.

For example, an image of the upper part of a user's body is captured by the camera module 54 through its lens. This is converted from an optical signal into an
10 electrical signal by the camera circuit 502 and output to the pre-processing circuit 503 as the digital image signal S502.

The image data of the upper part of the user's body is supplied to the display circuit 513 through the image
15 combining circuit 512. The image of the upper part of the user's body, that is, the initial image IIM, is then displayed on the display monitor 53.

While viewing the image of the upper part of his or her own body, that is, the initial image IIM, the user
20 decides on an initial reference image IRI by for example the first method, namely, monitoring the size frame of the image to be transmitted image.

That is, the transmission image size frame is set beforehand in the initial image IIM, and the image in the
25 frame FR is output as a monitor image MNTa and displayed

on the display monitor 53 by the display circuit 513.

Then, the user views the output on the monitor and when finding a desired image state, for example presses the switch 56 to issues a command for locking the image to the CPU 516 to thereby decide on the initial reference image IRI.

Then, when the initial reference image information is transmitted from the CPU 516, tracking is started in the pre-processing circuit 503.

10 In this state, in the input image data (the current image) CIM newly input from the camera circuit 502, a search region SRG is extracted by the search region extractor 5031 based on the coordinate information of the reference image RIM of the previous image PIM stored in
15 the reference memory 5034.

Next, the luminance signal of the extracted data is converted into binary data by the binary-coded unit 5033, then pattern matching of the luminance information is performed by the pattern matching unit 5033 between the
20 luminance signal of the extracted data and the binary data of the reference image RIM stored in the reference memory 5034. This processing result is stored in the reference memory 5034 as a reference image of the next image.

25 Then, the rectangular region surrounding the

tracking result MMT obtained with the pattern matching unit 5033 is cut off in the rectangular region extractor 5035 and it is output to the video compressor and coder 504 and the image combining circuit 512.

5 In such a state, the voice input from the microphone 501 is converted into digital and then is compressed by the voice compressor and coder 505.

Then, the image data obtained from the pre-processing circuit as described above is compressed and
10 coded by the image compressor and coder 504, multiplexed by the multiplexer 506 together with the compressed voice stream, and then output to the transmission circuit 507.

After that, the transmission circuit 507 receives the multiplexed data, the transmission protocol,
15 transmits the result for example to the mobile base station 61-1, and transmits it by the transmission channel 60 to for example the terminal 50-2 of the other party.

In addition, in the reverse order, the transmitted
20 signal from the other party is received by the reception circuit 508, demodulated, has its transmission protocol decompressed, then is separated into the image data stream and the voice data stream by the demultiplexer 509.

25 The separated voice data stream is decoded by the

voice decoding circuit 511 and sent to the speaker 514 for reproduction.

The image data stream is decoded by the video decoding circuit 510, then is combined with the user's own image from the pre-processing circuit 503 by the image combining circuit 512. The combined image data is output to the display circuit 513 and displayed on the display monitor 53.

At this time, the display monitor 53 displays the user's own face image and a transmitted face image of the other party are displayed for example side by side above the lens of the camera module 54 on the display monitor 53 and displays a specified content image is displayed below the lens.

As described above, according to the first embodiment, during communication between terminals 50-1 and 50-2, the target image is compressed and transmitted after being adjusted to fully fill the frame of the transmitted image, so the amount of information can be greatly reduced while transmitting the expression of the other party, which is needed most for the communication.

In addition, because the face image is adjusted to fully fill the frame, transmission of the high resolution background image can be naturally eliminated, and the amount of information after compression can be largely

reduced further. Consequently, it is possible to transmit an image of a high quality.

Further, the frame correlation of images can be increased greatly. By using a compression technique of
5 interframe compression like MPEG for the image compression method, the amount of information can be dramatically reduced and therefore it becomes possible to transmit an image of a high quality.

Further, the image state that a user desires to
10 transmit (image locking state) can be kept automatically. This helps detecting the characteristic portion of the image and enables tracking of the target image such as the image of the person. As a result, the time correlation of images is increased, and the amount of
15 information after compression can be decreased, and so it becomes possible to transmit an image of a high quality.

Further, since face images and content for common discussion are arranged at the upper and lower portions of the display monitor 53 and since a lens part of the
20 camera module 54 is provided between them, a state can be realized where the speakers are in better eye contact.

Further, the state of the users talking while viewing an image for discussion can be transmitted. Further, there is an advantage that virtual reality
25 communication where the users maintain eye contact by

looking at each others' faces on the monitor can be realized.

Furthermore, in the transmission channel 60, an application service provider (ASP) 66 able to distribute content may be started up and the same images of content 5 CONT provided by the application service provider 66 may be output to the terminals 50-1 and 50-2 through the gateway 65.

Second Embodiment

10 Figure 9 is a circuit diagram of a second embodiment of the signal processing system of a terminal functioning as a data transmission apparatus according to the present invention.

The point of difference of the signal processing 15 system of a terminal according to the second embodiment and the aforesaid signal processing system of according to the first embodiment lies in the processing of the pre-processing circuit 503A and the CPU 516A.

Specifically, during communication, a part of the 20 face is detected in the input image, the size of the face is calculated from that information, the image is enlarged or reduced so that the face image fully fills the monitor, then the face image is cut out.

In this case, the image of the characteristic 25 portion of the image to be transmitted is used to

calculate the distance of the characteristic portion and this used for enlarging, reducing, or tracking the image.

Particularly, in case of a face image, for example, the portion of the eyes is taken out for the

5 characteristic portion, the rate of enlargement or reduction of the image is determined from the distance between the two eyes so that the distance becomes the same as that in a reference image, then the image is tracked. In addition, as the characteristic portion, use

10 may be made of at least part of the eyes, nose, mouth, ears, or outline of the face and the image enlarged, reduced, or tracked from their sizes and distances.

The rest of the configuration is the same as that shown in Fig. 4. Therefore, in Fig. 9, components the

15 same as in Fig. 4 are shown with the same reference numerals.

Namely, the signal processing system 500A of a terminal, as shown in Fig. 9, has a microphone 501, a camera circuit 502, a pre-processing circuit 503A, a

20 video compressor and coder 504, a voice compressor and coder 505, a multiplexer (MUX) 506, a transmission circuit 507, a reception circuit 508, a demultiplexer (DeMUX) 509, a video decoding circuit 510, a voice decoding circuit 511, an image combining apparatus 512, a

25 display circuit 513, a speaker 514, a GUI interface (I/F)

circuit 515, and a CPU 516A.

Because the functions of the pre-processing circuit 503A and the CPU 516A are different from those in the first embodiment, below, only the functions of the pre-processing circuit 503A and the CPU 516A will be described in detail. Descriptions of other components will be omitted.

Figure 10 is a circuit diagram of an example of the specific configuration of the principal portion of the pre-processing circuit according to the second embodiment.

For easy understanding, the CPU 516A is also illustrated in Fig. 10.

The pre-processing circuit 503A, as shown in Fig. 10, has a face part position detector 5036, an image enlarging and reducing unit 5037, an image extractor 5038, and a reference face image memory 5039.

The face part position detector 5036 detects the position of parts of the face such as the two eyes, nose, mouth, eyebrows, ears, and the outline of the face within the information on the initial range of the face position by the CPU 516A, that is, a face part search region, by using information on luminance and colors and supplies the detected face part position information S5036 to the CPU 516A.

The image enlarging and reducing unit 5037 enlarges or reduces the input image based on the rate of enlargement or reduction calculated by the CPU 516A and supplies the same to the image extractor 5038.

- 5 The image extractor 5038 cuts out the face image in the region displayed on the monitor screen from the enlarged or reduced image based on the face position information (coordinates) already sent from the CPU 516A.

- 10 The cut out face image is sent to the reference face image memory 5039. At this time, a reference image is made from the rate of enlargement or reduction of the image.

- 15 Next, a description will be given of the operation of the pre-processing circuit 503A, including the processing of the CPU 516A with reference to Fig. 11 to Fig. 16.

- 20 Figure 11 is a view of an example for explaining the calculation of a face size, Fig. 12 is a view of an example for explaining the search region for face part detection, Fig. 13 and Fig. 14 are views for explaining processing for reducing and displaying a face image in a search region for face part detection on a display monitor, and Fig. 15 and Fig. 16 are views for explaining processing for enlarging and displaying a face image in a search region for face part detection on a display
- 25

monitor.

First, for the first input image (the initial image), information on the initial region of the face position is supplied to the CPU 516A by instructions by a user through the GUI interface circuit 515 (switch 56).

In the CPU 516A, the transmitted initial region is supplied to the face part position detector 5036 as a face part detection search region signal S516a.

In the face part position detector 5036, the position of parts of the face such as the two eyes, nose, mouth, eyebrows, ears, and outline of the face is detected in the supplied face part search region by using information on luminance and colors. Then, the position information on the face parts detected in the face part detector 5036 is supplied to the CPU 516A as a signal S5036.

In the CPU 516A, the size of the face in the image is calculated based on the position information of the face parts.

For example, as shown in Fig. 11, if denoting the distance between the detected two eyes as "a" and the perpendicular distance between the eyes and the mouth as "b", the size of the face is set to 3a in the horizontal direction and 4b in the vertical direction.

Note that in this case, it is also possible to

detect the color of the face to extract the contour and calculate the size of the face from that size.

Further, in the CPU 516A, the face part detection search region of the next image is calculated based on
5 the position information of the face parts.

For example, as shown in Fig. 12, assume the region of $\pm a \times N$ horizontally and $+b \times M$ and $-b \times L$ vertically around the rectangle $a \times b$ including the two eyes and the mouth is the face part detection search region of the
10 next image.

Note that the search region cannot be larger than the input image.

From the second input image and on, the face part detection search region obtained in the previous image is
15 supplied to the face part position detector 5036 by the CPU 516A.

In the face part position detector 5036, the position of face parts is detected in the same way as above in the supplied search region.

At this time, information of the face portion of the previous image stored in the reference face image memory 5039 is used as a reference image for detection of the face parts.
20

Then, the position information on face parts
25 detected in the face part detector 5036 is supplied to

the CPU 516A as a signal S5036.

Then, in the same way as the initial image, the size of the face in the image and the face part detection search region of the next image are calculated in the CPU 516A.

Figure 13 and Fig. 15 show the search regions of face parts.

Next, in the CPU 516A, the rate of enlargement or reduction is calculated from the calculated size of the face.

This rate is that for enlarging or reducing the face so that the face image just fully fills the screen of the monitor on which the user's face is displayed.

For example, as shown in Fig. 14, since the size of the face is $3a$ in the horizontal direction and the size of the screen is d in the horizontal direction, the reduction rate $d/3$ is obtained.

In addition, in Fig. 16, the enlargement rate $d/3a'$ is obtained.

The rate calculated by the CPU 516A is supplied to the image enlarging and reducing unit 5037 as a signal S516c.

In the image enlarging and reducing unit 5037, the input image is enlarged or reduced based on the signal S516c.

The enlarged or reduced image is supplied to the image extractor 5038. In the image extractor 5038, the face image in the region displayed on the monitor screen from the enlarged or reduced image is cut out based on
5 the face position information (coordinates) that has already been sent from the CPU 516A as the signal S516b.

Figure 14 and Fig. 16 show the cases in which the face size in the input image is larger than the monitor screen and the face size is smaller than the monitor
10 screen, respectively.

Although the sizes of the faces in the input images are different, the sizes of the face images displayed on the monitor screen are the same.

The cut out face image is supplied to the video
15 compressor and coder 504 and the later image combining circuit 52.

Further, the cut out face image is sent to the reference face image memory 5039. At this time, a reference image is constructed from this image and the
20 rate of enlargement or reduction.

According to the second embodiment, similar effects as the first embodiment can be obtained.

Third Embodiment

Figure 17 is a circuit diagram of a third embodiment
25 of the signal processing system of a terminal functioning

as a data transmission apparatus according to the present invention.

Figure 18 is a circuit diagram of an example of the specific configuration of the principal portion of the pre-processing circuit according to the third embodiment.

The point of difference between the signal processing system of a terminal according to the third embodiment and the aforesaid signal processing system according to the second embodiment lies in the processing of the pre-processing circuit 503B and the CPU 516B.

Specifically, the face image is enlarged or reduced and the face image is cut out so that the face image is constantly maintained to a size specified by the user.

For this, the pre-processing circuit 503B is configured to have add a monitor/face image ratio memory 5040 to the circuit in Fig. 10.

The rest of the configuration is the same as that shown in Fig. 9 and Fig. 10. Therefore, in Fig. 17 and Fig. 18, components the same as those in Fig. 9 and Fig. 10 are shown by the same reference numerals.

Namely, the signal processing system 500B of a terminal, as shown in Fig. 17, has a microphone 501, a camera circuit 502, a pre-processing circuit 503B, a video compressor and coder 504, a voice compressor and coder 505, a multiplexer (MUX) 506, a transmission

circuit 507, a reception circuit 508, a demultiplexer (DeMUX) 509, a video decoding circuit 510, a voice decoding circuit 511, an image combining apparatus 512, a display circuit 513, a speaker 514, a GUI interface (I/F) circuit 515, and a CPU 516B.

The pre-processing circuit 503B, as shown in Fig. 18, has a face part position detector 5036, an image enlarging and reducing unit 5037, an image extractor 5038, a reference face image memory 5039, and a monitor/face image ratio memory 5040.

Because the functions of the pre-processing circuit 503B and the CPU 516B are different from those in the second embodiment, below, only the functions of the pre-processing circuit 503B and the CPU 516B will be described in detail. Descriptions of other components will be omitted.

The monitor/face image ratio memory 5040 stores values for calculating the size of the face to be displayed on the monitor screen and the position of the face in the monitor screen based on the position information on face parts extracted by the CPU 516B. The stored data is used as a reference value in the second input image and on.

Next, the operation of the pre-processing circuit

503B, including the processing of the CPU 516B, will be described with reference to Fig. 19 to Fig. 23.

Figure 19 and Fig. 20 are views for explaining processing for reducing a face image in a specified image transmission region and displaying it on a display monitor, Figure 21 is a view for explaining the ratio of the pixels "e" of face color over the pixels "f" in the image on the monitor, and Fig. 22 and Fig. 23 are views for explaining processing for reducing a face image in a face part detection search region and displaying it on a display monitor.

First, the user specifies the region of his or her own face desired to be displayed on the monitor (desired to be transmitted) in the first input image (the initial image) on for example the GUI of the display monitor. This information on the region is supplied to the CPU 516B.

The region is easy for the user to specify is resembling the monitor.

In the CPU 516B, the supplied initial region is supplied to the face part position detector 5036 as a face part detection search region signal S516a.

In the face part position detector 5036, the position of parts of the face such as the two eyes, nose, mouth, eyebrows, ears, and outline of the face is

detected in the supplied face part search region. The position information on face parts detected in the face part detector 5036 is supplied to the CPU 516B as a signal S5036.

5 In the CPU 516B, the size of the face displayed on the monitor screen and values for calculating the position of the face in the monitor are extracted based on the position information of face parts.

For example, in the first input image (the initial
10 image), as shown in Fig. 20, the horizontal length "c" of the region specified by the user, the distance "a" between the detected two eyes in the initial image, and the horizontal length "d" of the display monitor 53 can be used.

15 Further, as shown in Fig. 21, it is possible to detect the face color and calculate the ratio of the pixels "e" of face color over the pixels "f" in the image on the monitor.

These values are sent to the ratio memory 5040 and
20 used as reference values from the second input image and on.

Note these values may also be stored in the CPU 516B.

Further, in the CPU 516B, the face part detection
25 search region of the next image is calculated based on

the position information of the face parts.

For example, in the same way as the second embodiment, as shown in Fig. 12, assume the region of $\pm a \times N$ horizontally and $+b \times M$ and $-b \times L$ vertically around
5 the rectangle $a \times b$ including the two eyes and the mouth is the face part detection search region of the next image.

For the second input image and on, the face part detection search region obtained in the previous image is
10 supplied to the face part position detector 5036 by the CPU 516B.

In the face part position detector 5036, positions of parts of the face are detected in the supplied search region in the same way as the above.

At this time, information of the face portion of the previous image stored in the reference face image memory
15 5039 is used as a reference image for detection of parts of the face.

The position information on face parts detected in
20 the face part detector 5036 is supplied to the CPU 516B as a signal S5036.

Next, in the CPU 516B, the rate of enlargement or reduction is calculated from the calculated size of the face.

In the first image (initial image), for example, the reduction rate d/c is obtained from the horizontal length "c: of the region specified by the user and the horizontal length "d: of the screen stored in the ratio memory 5040.

For the second image and on, the ratio is calculated from the face position information sent from the face part position detector 5036 and the reference values stored in the ratio memory 5040. For example, as shown in Fig. 23, the reduction rate a'/a is calculated from the distance a' between the two eyes sent from the face part position detector 5036 and the distance "a" between the two eyes in the first image stored in the ratio memory 5040 as a reference value.

Further, the reduction rate may also be found using the skin color pixels "g" from the face part position detector 5036 and the skin color pixels "e" in the initial image stored in the ratio memory 5040.

The rate calculated by the CPU 516B is supplied to the image enlarging and reducing unit 5037 as a signal S516c.

In the image enlarging and reducing unit 5037, the input image is enlarged or reduced based on the signal S516c.

The enlarged or reduced image is supplied to the

image extractor 5038. In the image extractor 5038, the face image in the region displayed on the monitor screen from the enlarged or reduced image is cut out based on the face position information (coordinates) already sent
5 from the CPU 516B as the signal S516b.

Figure 23 shows the state where the size of the face in the input image which is larger than the monitor screen is reduced so as to be displayed at the same as the face size specified by the user in the initial image
10 and where the size of the face image displayed in the monitor screen becomes always the same.

When the face in the input image is smaller than the initial image, it will be enlarged.

The extracted face image is supplied to the later
15 video compressor and coder 504 and image combining circuit 52.

Further, the extracted face image is sent to the reference face image memory 5039. At this time, a reference image is constructed from this image and the
20 rate of enlargement or reduction.

According to the third embodiment, similar effects as the first embodiment can be obtained.

Fourth Embodiment

Next, a specific example of improving the tracking
25 accuracy of an image by fetching and tracking the image

at a frame rate the same as or higher than the image to be transmitted and refreshing a reference image serving as a reference for tracking by a rate the same as or higher than the transmission frame rate will be explained
5 as a fourth embodiment of the present invention.

Figure 24 is a conceptual view for explaining the processing for improving the tracking accuracy of an image by fetching and tracking the image at a frame rate the same as or higher than the image to be transmitted
10 and refreshing a reference image serving as a reference for tracking by a rate the same as or higher than the transmission frame rate

Note that Fig. 24 shows the case in which the frame rate of the image to be transmitted is one-third the
15 input frame rate. In practice, the frame rate ratio is controlled according to the bandwidth of the transmission channel and the congestion state and is not set constant.

In the example of Fig. 24, each input frame is refreshed using the face image of the previous image as a
20 reference image and matching is performed in the search region of the next image.

For example, a face image detected in 100-1 of Fig. 24 is used as the reference image 100-8 of the next image, and matching is performed for 100-8 in the image
25 100-2 to detect the face image. Then, the face image

detected here is used as the next reference image.

Further, from the position of this face image, the search region of the next input image is decided and refreshed.

- 5 In accordance with the frame rate of the image to be transmitted, the image to be transmitted is selected from the face image detected in the input frame.

In Fig. 24, face images cut out in images 100-1, 100-4, and 100-7 are transmitted.

- 10 When setting the frame rate of image transmission the same as that of reference image refreshing, the time interval between the reference image and the image to be matched becomes longer, therefore there is a possibility that the matching accuracy will become poor for example
- 15 if the face image becomes gradually smaller, the expression changes drastically, or the face is turned at a slant.

- For example, when matching 100-8 and 100-4, the sizes of the face images are different, so the matching
- 20 accuracy is poor. In addition, when setting the rate of the refreshing of the search region the same as the frame rate of the image to be transmitted, if the face is moved horizontally, sometimes the face image will leave the search region, so there is a possibility the matching
- 25 accuracy will become poor.

For example, when using the search region of 100-1 for matching 100-4, the face image will not be covered by the search region at all, so the matching accuracy will be poor.

5 As in the fourth embodiment, if the image is fetched and tracked at a frame rate the same as or higher than the image to be transmitted and a reference image serving as a reference for tracking is refreshed at a rate the same as or higher than the transmission frame rate, the
10 tracking accuracy can be improved.

According to the present invention, the total amount of information can be greatly reduced while transmitting the expression of the other party which is needed most for communication.

15 In addition, according to the present invention, because the face image is adjusted to fully fill the frame, transmission of the high resolution background image can be naturally eliminated, and the amount of information after compression can be largely reduced.
20 Consequently, it is possible to transmit an image of a high quality.

Summarizing the effects of the invention, according to the present invention, the frame correlation of images is able to be increased greatly. By using a compression
25 technique of interframe compression like MPEG for the

image compression method, the amount of information can be decreased drastically, and therefore it becomes possible to transmit an image of a high quality.

In addition, according to the present invention, the
5 image state that a user desires to transmit (image locking state) can be kept automatically. This helps detect the characteristic portion of the image and enables tracking of the target image such as the image of a person. As a result, the time correlation of images is
10 increased, and the amount of information after compression can be decreased. Therefore it becomes possible to transmit an image of a high quality.

Further, according to the present invention, a state can be realized in which the users maintain better eye
15 contact.

Further, according to the present invention, the state of the users talking while viewing an image for discussion can be transmitted. Further, there is an advantage that virtual reality communication where the
20 users maintain eye contact by looking at each others' faces on the monitor can be realized.

While the invention has been described with reference to specific embodiment chosen for purpose of illustration, it should be apparent that numerous
25 modification could be made thereto by those skilled in

the art without departing from the basic concept and
scope of the invention.

COPIES OF THIS DOCUMENT
ARE AVAILABLE FROM THE
NATIONAL ARCHIVES
AT COLLEGE PARK, MARYLAND
20740